Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 28 May 2026
Revised 22 June 2026
Accepted 17 July 2026

Natural Resources for Human Health 2026, Vol. 6 (9s): 625–633

KEYWORDS:
chronic kidney disease, CKD staging, deep neural network, artificial neural network, multi-stage classification, clinical prediction, laboratory parameters

Stage-wise Classification of Chronic Kidney Disease (Stages 1–5) using a Deep Artificial Neural Network on Routine Clinical Blood Parameters

Utkarsha Rane[1*], Dr. Satish Ramesh Kolhe[2]

[1,2] School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon, India

ABSTRACT: Chronic kidney disease (CKD) is often diagnosed late because early stages usually have no symptoms. Identifying the stage of CKD is important for guiding treatment and monitoring progression. In this study, we developed a deep neural network to classify patients into all five CKD stages using routine clinical and laboratory data. The model uses patient demographics, lab results, and comorbidity information to make predictions without invasive tests. The model is trained and evaluated on a single-center dataset comprising 500 anonymized patient records. On the test set, it achieved 98% overall accuracy, with high precision and recall for Stages 1 to 4. The dataset exhibits notable class imbalance, with advanced CKD stages being underrepresented, and Stage 5 performance is limited due to the very small number of available samples. The results show that early and intermediate stages can be identified reliably using standard data. This approach could help diagnostic laboratories support clinicians in decision-making and monitoring. Future work will focus on validation using larger, multi-center datasets and on improving model interpretability.

1. INTRODUCTION

Chronic kidney disease (CKD) is a progressive and irreversible condition that affects a growing proportion of the global population and has emerged as a major public health challenge. Recent estimates indicate that nearly 788 million individuals worldwide are living with CKD, making it one of the leading contributors to morbidity and mortality across many regions [1]. A key challenge in CKD management is that the disease often remains asymptomatic in its early stages, which frequently results in delayed diagnosis and intervention. Clinically, CKD is defined by persistent abnormalities in kidney structure or function lasting for more than three months and is classified into five stages based on estimated glomerular filtration rate (eGFR) and albuminuria levels, reflecting increasing severity of renal impairment [2], [3].

Accurate differentiation between CKD stages is essential for effective clinical management. Treatment strategies vary considerably across disease stages, with early stages typically managed through medication and lifestyle interventions, while advanced stages require intensive monitoring and preparation for renal replacement therapy. CKD is also strongly associated with cardiovascular complications, increased hospitalization rates, and reduced life expectancy, contributing to a substantial healthcare burden [4], [5]. Consequently, reliable stage-wise assessment plays a critical role in guiding treatment decisions, monitoring disease progression, and reducing the risk of progression to end-stage kidney disease (ESKD) [6].

Traditional diagnostic approaches and rule-based scoring systems remain widely used in clinical practice; however, they exhibit notable limitations when applied to CKD staging. Disease progression is influenced by a complex interplay of demographic, clinical, and biochemical factors, many of which exhibit non-linear relationships across disease stages. Fixed thresholds and linear statistical models often fail to capture these interactions adequately [7]. Moreover, rule-based systems rely on manually defined decision criteria, which limits their adaptability to high-dimensional clinical data and reduces their ability to generalize across patient populations with varying comorbidities and disease trajectories, particularly in electronic health record–based

datasets [8]. As a result, conventional approaches may provide an incomplete representation of disease severity in real-world clinical settings.

In recent years, machine learning (ML) techniques have gained increasing attention for CKD detection and prediction tasks. Supervised algorithms such as Support Vector Machines, Random Forests, and Logistic Regression have demonstrated strong performance in distinguishing CKD patients from healthy individuals using routine clinical and laboratory features [9], [10]. Despite their effectiveness, most existing ML-based studies focus primarily on binary classification or early CKD detection. Their ability to differentiate between intermediate and advanced CKD stages remains limited, which restricts their usefulness for detailed clinical decision-making and stage-specific intervention planning [11].

More recently, deep learning and hybrid architectures, including CNN–LSTM–GRU models, have been proposed to capture complex feature interactions and temporal patterns in clinical data. These approaches have reported improved classification accuracy compared to traditional ML models, highlighting the potential of deep representations for CKD-related tasks [12]. Nevertheless, fine-grained CKD staging using real-world laboratory data remains relatively underexplored [13], [14]. Most existing studies either employ coarse multi-class groupings or lack sufficient validation across clinically meaningful CKD stages. This reveals a clear research gap for robust, stage-aware models that can distinguish all five CKD stages using routinely collected clinical data.

In this study, we address this gap by proposing a deep artificial neural network (ANN) framework for comprehensive stage-wise classification of CKD. Unlike prior work that focuses on binary detection or limited staging, the proposed model is designed to distinguish all five CKD stages using only standard blood and clinical parameters commonly available in diagnostic laboratories. The model's performance is systematically evaluated using learning curves, class-wise metrics, and confusion matrix analysis, with particular attention to the impact of class imbalance on advanced-stage prediction. By providing a practical, data-driven approach to fine-grained CKD staging, this work aims to support more accurate disease stratification and informed clinical decision-making in real-world settings.

2. DATASET DESCRIPTION

2.1 Data Source

The study uses a manually curated clinical dataset comprising records of 500 patients diagnosed with chronic kidney disease. The data were collected from a nephrology clinic and dialysis center in Jalgaon, Maharashtra, India, as part of routine clinical diagnostic practice under the supervision of a licensed nephrologist. Data access was granted for research purposes, and all patient records were anonymized before analysis to ensure privacy and ethical compliance.

2.2 Feature Description

The dataset includes demographic, clinical, biochemical, and comorbidity information:

- Demographics: Age and Gender.
- Clinical and Biochemical Indicators: Blood Pressure, Albumin, Blood Glucose, Serum Creatinine, Blood Urea Nitrogen (BUN), Albumin-to-Creatinine Ratio (ACR), Urine Protein-to-Creatinine Ratio (UPCR), Hemoglobin, and Estimated Glomerular Filtration Rate (eGFR).
- Comorbidities: Presence of Diabetes and Hypertension.
- Target Variable: CKD stage, ranging from 1 to 5, representing increasing disease severity.

2.3 Data Characteristics

The dataset comprises 500 patient records with no missing values. Analysis of the CKD stage distribution reveals a significant class imbalance, with Stage 1 being the most frequent (157 samples) and Stage 5 extremely rare (10 samples). This imbalance highlights the challenge of accurately predicting advanced CKD stages and emphasizes the need for careful evaluation of model performance across all classes.

3. METHODOLOGY

3.1 Data Preprocessing

Before training the model, the dataset was preprocessed to ensure compatibility with the neural network and to improve learning efficiency. Categorical variables, including gender, diabetes, and hypertension, were converted into a numerical format using one-hot encoding. All features were then normalized using the StandardScaler to standardize their ranges, ensuring that variables such as blood pressure and serum creatinine contribute appropriately during training. The dataset was subsequently divided into training and testing sets using an 80:20 split, with the training set used to optimize the model parameters and the testing set reserved for evaluating performance.

3.2 Deep Neural Network Architecture

A deep ANN was constructed using the Keras API to classify patients into CKD stages. The architecture was designed to capture complex relationships among clinical and biochemical features while controlling for overfitting. The input layer contained several neurons corresponding to the processed features. This was followed by three fully connected hidden layers with 64, 32, and 16 neurons, each employing the ReLU activation function. Dropout regularization was applied to the first two hidden layers at rates of 30% and 20%, respectively, to prevent overfitting. The output layer consisted of five neurons with Softmax activation, producing the probability distribution across the five CKD stages.

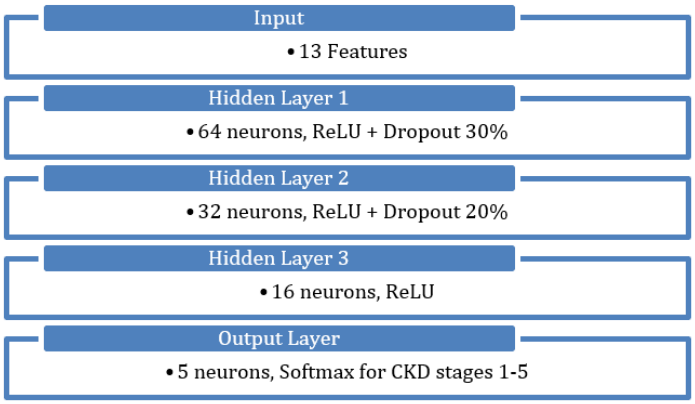

Fig 1. Deep ANN architecture for CKD stage prediction

3.3 Model Training

The network was trained using the categorical cross-entropy loss function, appropriate for multi-class classification tasks. The Adam optimizer updated network weights efficiently. Training was performed for 50 epochs with a batch size of 32 using a learning rate of 0.001. All experiments were conducted using Python in a standard workstation environment equipped with an Intel Core i5 (or equivalent AMD Ryzen 5) processor. Model development and training were performed using the Keras API with a TensorFlow backend. We performed data preprocessing and analysis using standard Python libraries, including NumPy, Pandas, and Scikit-learn.

4. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

4.1 Learning Curve Analysis

Fig. 2 shows the training and validation accuracy and loss curves obtained during model training. Accuracy improves rapidly in the initial epochs and begins to stabilize after around 25 epochs. Both training and validation accuracy converge to approximately 98%, with only minor fluctuations thereafter. The close alignment between the two curves suggests that the model learns consistently without noticeable overfitting.

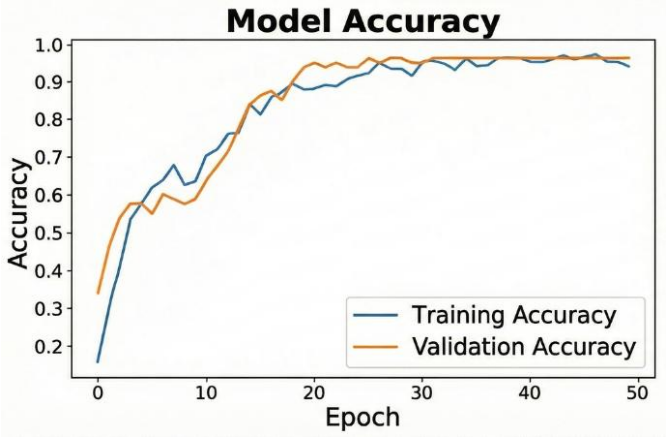

Fig 2. Model Accuracy

A similar trend is observed in the loss curves (Fig. 3). Training and validation loss decrease steadily as the number of epochs increases and converge to low values in the later stages of training. The absence of divergence between the loss curves indicates stable optimization and satisfactory generalization to unseen data. Overall, the learning curves confirm that the proposed ANN converges effectively under the selected training configuration.

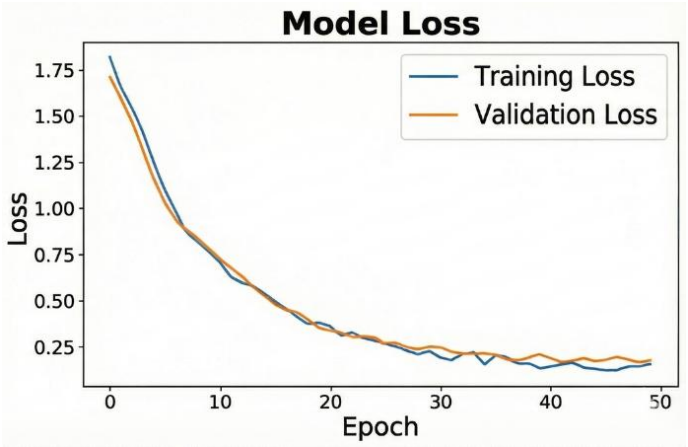

Fig 3. Model Loss

4.2 Classification Metrics

The trained model achieved an overall test accuracy of 98.00% with a corresponding test loss of 0.0815. These results indicate strong overall performance for multi-stage CKD classification.

Table I summarizes the class-wise precision, recall, and F1-score obtained on the test set. Stages 1 to 4 demonstrate consistently high performance, with F1-scores ranging from 0.97 to 1.00. Stage 3, in particular, is classified without error, achieving perfect precision and recall. Minor performance variation is observed for Stage 2, where a small number of samples are misclassified, leading to a slightly lower recall.

Table 1 Class-wise performance metrics for CKD stage classification

Stage	Precision	Recall	F1-Score	Support
Stage 1	0.97	1.00	0.98	32
Stage 2	1.00	0.96	0.98	23
Stage 3	1.00	1.00	1.00	28
Stage 4	0.94	1.00	0.97	16
Stage 5	0.00	0.00	0.00	1
Overall Accuracy			0.98	100

The reduced macro-level performance is primarily due to the poor results for Stage 5, which is severely underrepresented in the dataset.

4.3 Classification Outcome Statistics

To provide transparency on how performance metrics are derived, Table II summarizes the true positive (TP), false positive (FP), false negative (FN), and true negative (TN) counts for each CKD stage. These values form the basis for computing precision, recall, and F1-score.

Table 2 Classification outcome statistics for CKD stages

Stage	TP	FP	FN	TN
Stage 1	32	1	0	67
Stage 2	22	0	1	77
Stage 3	28	0	0	72
Stage 4	16	1	0	83
Stage 5	0	1	1	98

Precision, recall, and F1-score are computed using standard definitions based on these values, where precision = TP/(TP+FP), recall = TP/(TP+FN), and F1-score is the harmonic mean of precision and recall.

4.4 Confusion Matrix Interpretation

The confusion matrix shown in Fig. 4 further illustrates the classification behavior of the model. Most predictions lie along the diagonal, particularly for Stages 1 through 4, indicating reliable and consistent classification. Misclassifications are rare and generally occur between neighboring stages, which is expected given the gradual progression of CKD severity.

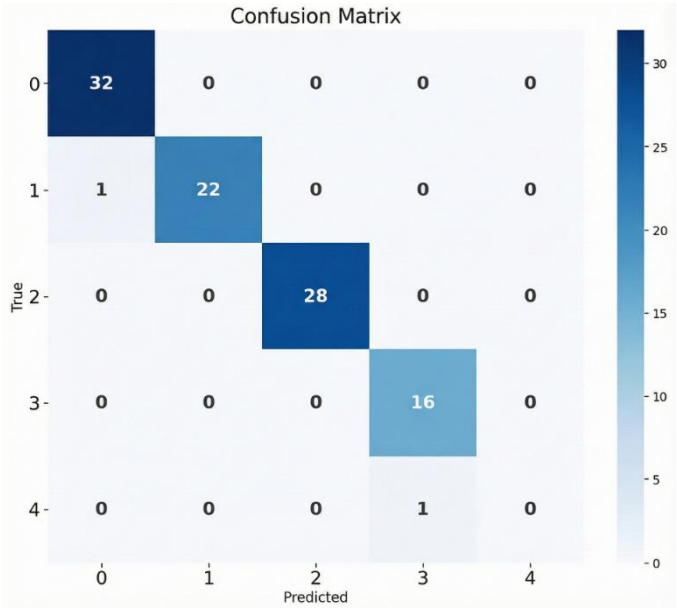

Fig 4. Confusion Matrix

Stage 5 shows incorrect classification, with the single test instance misclassified as Stage 4. This outcome reflects the extreme class imbalance rather than a limitation of the model architecture itself. The results emphasize the difficulty of learning robust decision boundaries for advanced CKD stages when very limited training data are available and highlight the need for larger, more balanced datasets in future studies.

5. DISCUSSION

The proposed model achieved an overall test accuracy of 98.00% with a corresponding loss of 0.0815, and the training and validation curves demonstrated stable convergence without noticeable overfitting. For CKD Stages 1 to 4, precision and recall remained consistently high, with F1-scores ranging from 0.97 to 1.00. In particular, Stage 3 was classified without error. These results indicate that routinely measured clinical and laboratory parameters are sufficient to reliably distinguish early and intermediate stages of CKD. From a clinical perspective, this is significant, as treatment strategies and monitoring protocols differ across CKD stages, and accurate early stratification can influence disease progression and complication risk.

Similar performance trends have been reported in prior machine-learning-based CKD studies. TabNet-based models trained on longitudinal clinical data have achieved multi-stage prediction accuracies above 91% while relying on standard variables such as age and serum creatinine [15]. Other supervised and ensemble learning approaches have also demonstrated strong performance, in some cases approaching 98% accuracy, using only routine laboratory and clinical features [16], [17]. These findings reinforce that complex or invasive inputs are not essential for effective CKD staging.

To further contextualize the performance of the proposed ANN, a comparison with baseline models reported in the existing literature is summarized in Table III. While many prior studies focus on binary CKD detection or coarse multi-class grouping, the proposed model explicitly addresses fine-grained stage-wise classification across all five CKD stages and achieves competitive overall accuracy under real-world data constraints.

The principal limitation of the proposed model lies in Stage 5 prediction. Only one Stage 5 sample was present in the test set, and it was misclassified as Stage 4. This behavior is evident in the confusion matrix and is primarily a consequence of severe class imbalance rather than a limitation of the network architecture. Similar challenges for minority CKD stages have been reported in earlier studies, where rare classes exhibit reduced sensitivity unless imbalance-aware strategies are employed [13].

Table 3. Comparison With Baseline CKD Classification Models Reported in Literature

Study	Task	Models Evaluated	Reported Accuracy
[18]	Multiclass CKD staging (Stages 1–5)	PNN, MLP, SVM, RBF	96.7%
[19]	Five-class CKD classification	SVM, RF, DT, XGBoost	82.56%
[20]	Multiclass CKD staging	RF, XGBoost, MLP	97.92%
This work	Stage-wise CKD classification (Stages 1–5)	Deep ANN	**98.00%**

Although imbalance-handling techniques such as synthetic oversampling, class-weighted loss functions, or focal loss are commonly used in multi-class learning, they were not applied in this study. Given the extremely limited number of Stage 5 samples, the use of such techniques would likely introduce artificial bias or unstable decision boundaries without guaranteeing meaningful generalization. Instead, the study prioritizes transparent reporting of performance under the original data distribution. Future work will investigate imbalance-aware training strategies when larger and more representative datasets for advanced CKD stages become available.

The dataset used in this study comprised 500 patient records collected from a single clinical center. While sufficient for evaluating early and moderate CKD stages, it does not fully capture the variability associated with advanced disease across broader populations. The limited representation of Stage 5 cases restricts conclusions regarding severe CKD prediction and highlights the need for larger, multi-center datasets to improve robustness and generalizability.

Model interpretability is also a key consideration for clinical adoption. Previous studies have shown that explainability techniques, such as SHAP-based feature attribution and attention-driven analysis, can help identify influential clinical parameters and align model outputs with medical reasoning in CKD prediction tasks [15], [21]. Incorporating such explainability tools would enhance clinician trust, support decision-making, and facilitate integration into clinical decision-support systems. This represents an important direction for future work. In particular, future multi-center validation studies will explicitly incorporate model interpretability analyses, including SHAP-based feature attribution, to systematically evaluate prediction rationale and ensure clinically meaningful transparency of stage-wise CKD classification outcomes.

6. CONCLUSION

In this study, we proposed a deep ANN framework for multi-stage CKD classification using routinely available clinical and laboratory parameters. The model is designed to capture complex relationships among demographic, biochemical, and comorbidity features, providing stage-aware predictions without relying on invasive or specialized tests.

Experimental results demonstrate strong overall performance, with an accuracy of 98.00% and stable convergence of training and validation curves. Class-wise evaluation shows consistently high precision, recall, and F1-scores for Stages 1 to 4, with Stage 3 classified without error. The primary limitation lies in Stage 5 prediction, which is constrained by severe class imbalance and the scarcity of advanced-stage samples. These findings highlight the robustness of the proposed model for early and intermediate CKD stages, while emphasizing the need for larger datasets to improve the prediction of rare, advanced stages.

Given its reliance on standard clinical and laboratory data, the proposed ANN framework has significant potential for real-world deployment in diagnostic laboratories. Its ability to provide accurate, stage-wise CKD assessment can support informed clinical decision-making, guide treatment strategies, and facilitate timely intervention to reduce the risk of progression to end-stage kidney disease. Future work will focus on integrating interpretability methods, with forthcoming multi-center validation studies explicitly incorporating SHAP-based analyses to enhance transparency, clinician trust, and clinical applicability of the proposed framework.

AUTHOR CONTRIBUTIONS

Conceptualization, U.R. and S.R.K.; methodology, U.R.; software, U.R.; validation, U.R. and S.R.K.; formal analysis, U.R.; investigation, U.R.; resources, U.R. and S.R.K.; data curation, U.R.; writing—original draft preparation, U.R.; writing—review

and editing, U.R. and S.R.K.; visualization, U.R.; supervision, S.R.K.; project administration, S.R.K.; funding acquisition, S.R.K. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

The dataset used in this study consists of anonymized clinical records of 500 chronic kidney disease patients collected from a collaborating nephrology clinic in Jalgaon, India, under the supervision of a licensed nephrologist. All personally identifiable information was removed before analysis to ensure patient privacy and compliance with ethical research practices. Although the data have been anonymized, the dataset originates from a clinical healthcare provider and is subject to institutional data-sharing restrictions. Therefore, it cannot be publicly released. However, the anonymized dataset may be made available for academic research purposes upon reasonable request to the corresponding author, subject to approval from the data provider.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENT

This research was not funded by any grant

REFERENCES

[1] Mark, P. B., et al. (2025). Global, regional, and national burden of chronic kidney disease in adults, 1990–2023, and its attributable risk factors: A systematic analysis for the Global Burden of Disease Study 2023. The Lancet, 406(10518), 2461–2482. https://doi.org/10.1016/S0140-6736(25)01853-7

[2] Qin, K., Qing, J., Wang, Q., & Li, Y. (2024). Epidemiological shifts in chronic kidney disease: A 30-year global and regional assessment. BMC Public Health, 24(1), 3519. https://doi.org/10.1186/S12889-024-21065-9

[3] Vaidya, S. R., & Aeddula, N. R. (2024). Chronic kidney disease. In Science basis of urology (pp. 257–264). https://doi.org/10.51249/hs.v4i01.1865

[4] Duff, R., et al. (2024). Global health inequalities of chronic kidney disease: A meta-analysis. Nephrology Dialysis Transplantation, 39(10), 1692–1709. https://doi.org/10.1093/NDT/GFAE048

[5] Okpechi, I. G., et al. (2022). Early identification of CKD—A scoping review of the global populations. Kidney International Reports, 7(6), 1341–1353. https://doi.org/10.1016/j.ekir.2022.03.031

[6] Alexiuk, M., & Tangri, N. (2024). Prediction models for earlier stages of chronic kidney disease. Current Opinion in Nephrology and Hypertension, 33(3), 325–330. https://doi.org/10.1097/MNH.0000000000000981

[7] Iliyas, I. I., Boukari, S., & Gital, A. Y. (2025). Recent trends in prediction of chronic kidney disease using different learning approaches: A systematic literature review. Journal of Medical Artificial Intelligence, 8. https://doi.org/10.21037/JMAI-24-256/COIF

[8] Mondal, R. S., & Bhuiyan, M. N. A. (2024). Predictive analytics for chronic disease management: A machine learning approach to early intervention and personalised treatment. Journal of Computer and Analytical Applications, 33(8), 4096–4107.

[9] Srilakshmi, V., et al. (2025). Chronic kidney disease prediction based on machine learning algorithms. International Journal of Information Technology and Computer Engineering, 13(2), 1033–1045. https://doi.org/10.62643/

[10] Gore, S. (2023). Brain tumour segmentation and analysis using BraTS dataset with the help of improvised 2D and 3D UNet model. Research Square. https://doi.org/10.21203/rs.3.rs-2791706/v1

[11] Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. Journal of Big Data, 9(1), 109. https://doi.org/10.1186/s40537-022-00657-5

[12] Fahim, A., Osman, A. M., Tarek, Z., & Elshewey, A. M. (2025). A hybrid CNN–LSTM–GRU deep learning model for the accurate classification of chronic kidney disease. Engineering, Technology & Applied Science Research, 15(6), 30657–30662. https://doi.org/10.48084/ETASR.14206

[13] Kumar Bajpai, A., Gudla, B., Mannanuddin, K., Ravi, K., & Singha, R. (2025). Machine learning models for prediction and classification in chronic kidney disease. South Eastern European Journal of Public Health, 3762–3769. https://doi.org/10.70135/SEEJPH.VI.3556

[14] Gore, S., Jadhav, S., Gore, S., Deshmukh, K. D., Rane, N., & Rao, P. (2026). AI for healthcare administration: Streamlining operations and patient management. In Lecture Notes in Networks and Systems (Vol. 1600, pp. 312–325). Springer. https://doi.org/10.1007/978-3-032-03072-6_25

[15] Chowdhury, M. N. H., et al. (2025). Deep learning for early detection of chronic kidney disease stages in diabetes patients: A TabNet approach. Artificial Intelligence in Medicine, 166, 103153. https://doi.org/10.1016/j.artmed.2025.103153

[16] Iftikhar, H., Hashem, A. F., Qureshi, M., & Rodrigues, P. C. (2025). Clinical application of machine learning models for early-stage chronic kidney disease detection. Diagnostics, 15(20). https://doi.org/10.3390/diagnostics15202610

[17] Metherall, B., Berryman, A. K., & Brennan, G. S. (2025). Machine learning for classifying chronic kidney disease and predicting creatinine levels using at-home measurements. Scientific Reports, 15(1), 4364. https://doi.org/10.1038/s41598-025-88631-y

[18] Rady, E. H. A., & Anwar, A. S. (2019). Prediction of kidney disease stages using data mining algorithms. Informatics in Medicine Unlocked, 15, 100178. https://doi.org/10.1016/j.imu.2019.100178

[19] Debal, D. A., & Sitote, T. M. (2022). Chronic kidney disease prediction using machine learning techniques. Journal of Big Data, 9(1), 109. https://doi.org/10.1186/s40537-022-00657-5

[20] Gogoi, P., & Valan, J. A. (2026). A multiclass machine learning framework for chronic kidney disease staging using CTGAN-based synthetic data augmentation and explainable AI. Computer Methods in Biomechanics and Biomedical Engineering, 1–18. https://doi.org/10.1080/10255842.2025.2610677

[21] Haque, M. E., Islam, S. M. J., Maliha, J., Sumon, M. S. H., Sharmin, R., & Rokoni, S. (2025). Improving chronic kidney disease detection efficiency: Fine tuned CatBoost and nature-inspired algorithms with explainable AI. In 2025 IEEE 14th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 811–818). IEEE. https://doi.org/10.1109/CSNT64827.2025.10968421